Subjects with mild COVID-19 and healthy control subjects

Subject ID	Age	Sex	DP SO	IgG	IgM	PNA	AR DS	Abnormal X-ray
201657034	20-25	M	24	1.45	1.07	No	No	No
201657036	30-35	M	33	6.60	2.20	No	No	No
201657037	30-35	F	27	36.9	< 1	No	No	No
201657783	55-60	M	24	> 15	< 1	No	No	No
201657785	55-60	F	25	> 15	< 1	No	No	Unknown
201657787	35-40	F	37	29.8	< 1	No	Yes	Yes
201657790	20-25	F	31	12.9	< 1	Yes	Yes	Yes
Mean	**38**		**29**					
2010113866	50-55	M	**Healthy controls**					
2010113873	30-35	F						
2010113875	25-30	M						
2010113904	20-25	M						
Mean	**36**							

Patients with severe COVID-19

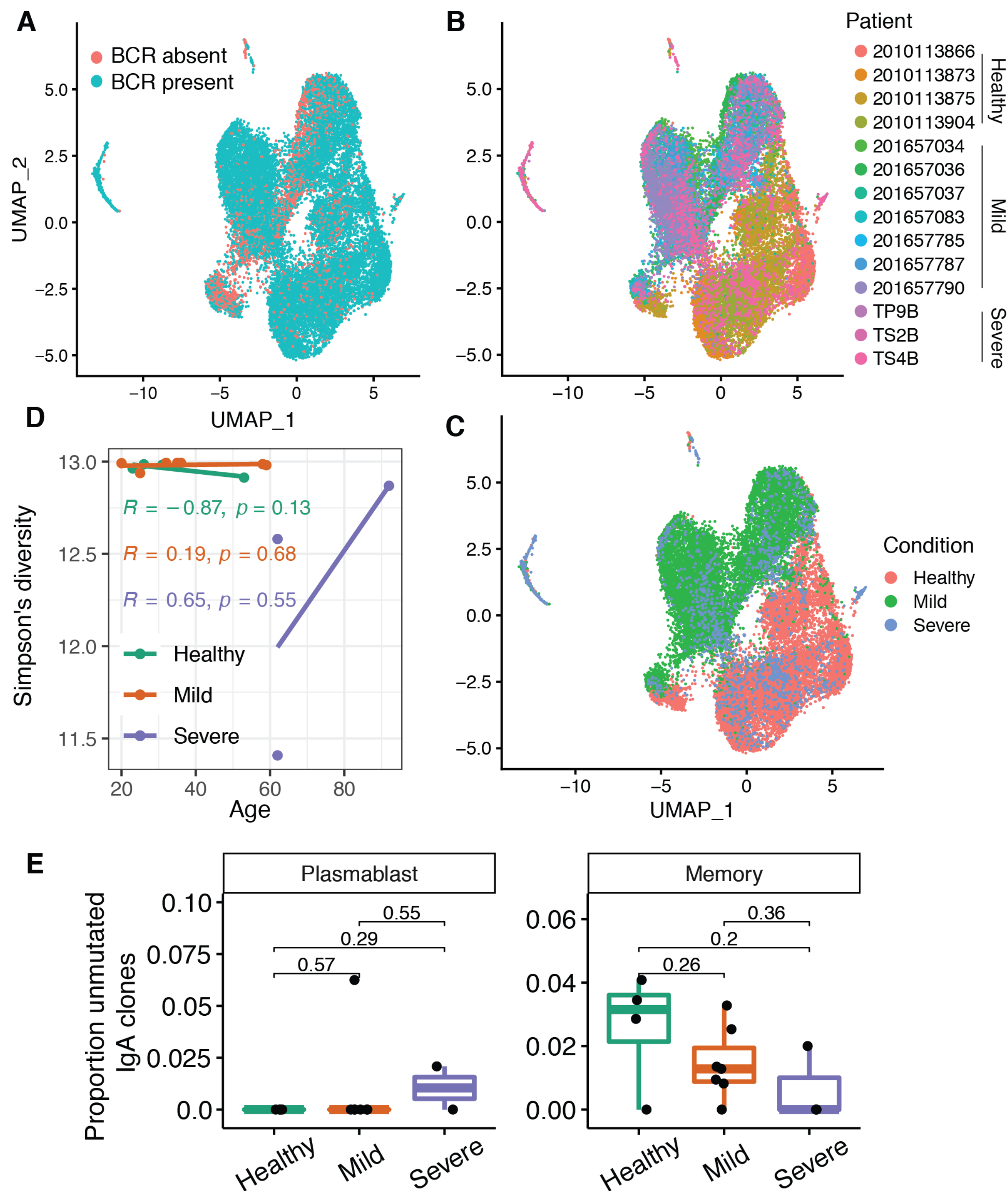
Patient ID	Age	Sex	DP SO	Status	TCZ	ATV	MP
TS2B	60-65	M	20	Stable	Yes	Yes	No
TS4B	90-95	M	23	Stable	Yes	Yes	No
TP9B	60-65	M	20	Prog.	Yes	Yes	Yes
Mean	**72**		**21**				

Supplemental Table I: Patient cohort information. IgG and IgM measurements are in (AU/mL), DPSO indicates days post symptom onset, PNA indicates pneumonia, and ARDS indicates acute respiratory distress syndrome. Abnormal X-ray indicates an abnormal chest X-ray. Treatment listed for patients with severe COVID-19: TCZ indicates Tocilizumab, ATV indicates Atazanavir, MP indicates Methylprednisolone.

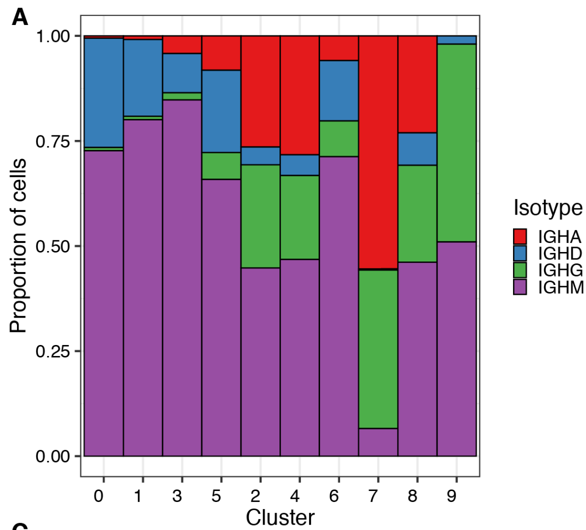
Supplemental Figure 1 (description): Detailed computational methods

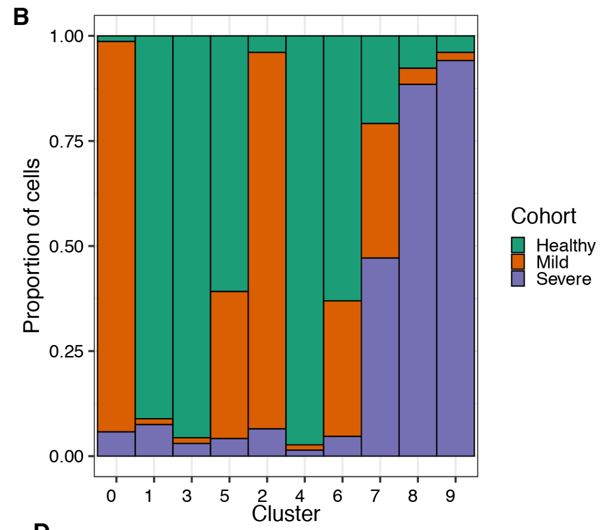
B cell subtype identification. Single cell RNA-seq gene expression information from subjects with mild COVID-19 and healthy controls was processed using Seurat v3.2.2 (12). To remove apoptotic or lysing cells, cells with a >10% of RNA transcripts from mitochondrial genes were excluded. To exclude poor quality cells, cells with reads from < 400 features were also removed. These cutoffs were determined by manual inspection of feature distribution plots. Read counts were log-normalized using a scaling factor of 10^4. To account for variability in gene expression, log-normalized read counts were then scaled and centered for each feature. The top 2000 variable genes were then identified using the "vst" method in Seurat. V(D)J genes from the IGH and IGL and IGK locus were removed so that the properties of the BCR expressed by the cell would remain independent of the cluster it is assigned to. This data was then reduced to the first 20 principal components. Cells were then clustered by the shared nearest neighbor clustering algorithm with a resolution of 0.25. The B cell subtype of each cluster was then determined by gene expression correlations to cell types in the immunoStates database (13). B cell subtype assignments were confirmed using known marker genes for plasmablasts (*XBP1*, *PRDM1*), memory B cells (*CD24*, *TNFRSF13B*) and naive B cells (*IGHD*, *IL4R*, *TCL1A*). B cell subtype assignments in patients with severe COVID-19 were previously performed as part of (9) (posted on preprint server). We filtered cells from each cohort to include only B cells. We then used the IntegrateData function of Seurat to combine the two datasets, and repeated clustering and cell type assignment with the integrated data as detailed previously. These clusters are visualized in **Fig. 1A**. Cluster assignments were confirmed by manual inspection of dot plots showing expression of marker genes (**Fig. 1B**), as well as SHM and isotype frequencies (**Supplemental Fig. 3D**). A small cluster of 51 cells (cluster 9) was identified as T/B cell doublets based on BCR and CD3E expression. Another cluster of 26 cells (cluster 8) was identified as monocyte/B cell doublets based on CD14 expression. These two clusters were retained during clonal clustering but removed afterwards. Annotations from patients with severe COVID-19 were concordant (1132/1186 cells) with prior annotations (9).

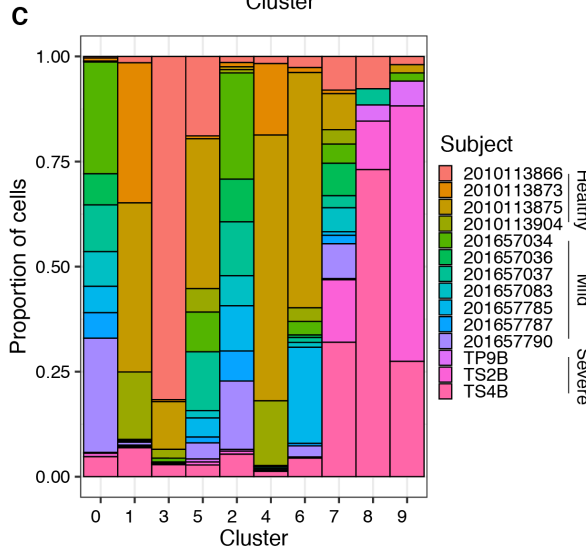
B cell receptor sequence processing and analysis. To obtain V, D, and J gene assignments, filtered contigs from CellRanger were aligned to the IMGT v3.1.24 germline reference allele database (14). Non-productive heavy and light chain BCR sequences were removed. BCRs were annotated by B cell subtype based on matching their single cell barcodes to the gene expression information. Only cells that had both gene expression and BCR sequence data were retained. In cells with multiple heavy chains, only one heavy chain with the highest unique molecular identifier count was retained. In the event of a tie, the first heavy chain identified was retained. To infer clonal clusters, sequences were partitioned based on common V and J gene annotations, as well as junction region length. Within these groups, sequences differing from one another by a Hamming distance threshold of 0.15 within the junction region were clustered into clones using single linkage hierarchical clustering (16). The Hamming distance threshold was determined by manual inspection of the distance to the nearest sequence neighbor plot using SHazaM v1.0.2 (17). Clonal clustering was performed using Change-O v1.0.0 (18). Clonal clusters were further separated if they contained light chains with differing V and J genes. Unmutated germline V and J gene sequences were reconstructed for each clone using Change-O. SHM level was determined for each cell as the frequency of non-ambiguous mismatches along the V gene of the unmutated germline sequence. Clonal diversity was performed using Alakazam v1.0.2.999 (18) and was calculated by taking the mean of 1000 re-sampling realizations to a uniform sequencing depth.

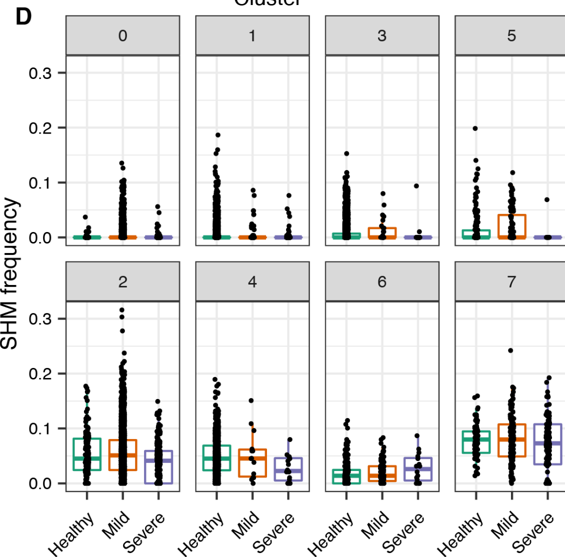


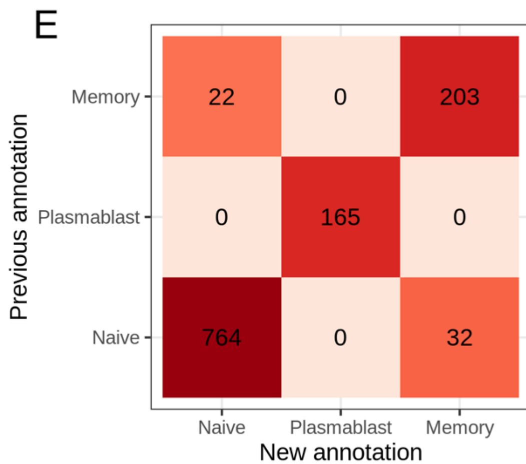
Supplemental Figure 2: UMAPs with cells colored by **A)** presence of BCR sequence, **B)** patient/subject ID, and **C)** patient cohort. **D)** Linear regression of Simpson's diversity of memory B cells vs age in each cohort (see **Fig. 2B**). R value shows correlation, p value shows lineage regression slope p value. **E)** Proportion of IgA plasmablast and memory B cell clones that are unmutated (median SHM < 1%) in each cohort. P values are from a Wilcoxon test. **Fig. 2C** shows similar plots for IgG clones.











Supplemental Figure 3: **A-D**) Isotype frequency, cohort frequency, subject frequency, and SHM frequency of cells within each cluster shown in **Fig. 1A**. **E)** Concordance of B cell type annotations between those used previously in (9) and new clustered annotations done in this study (New Annotation). Blocks within the table show the number of B cells in each category. Blocks are colored by the $\log_{10}$(cell count + 1). Diagonal blocks show concordant annotations, while off-diagonal blocks show discordant annotations.